Digital Repositories

Babak Hamidzadeh

Office of Strategic Initiatives

(October 2005)

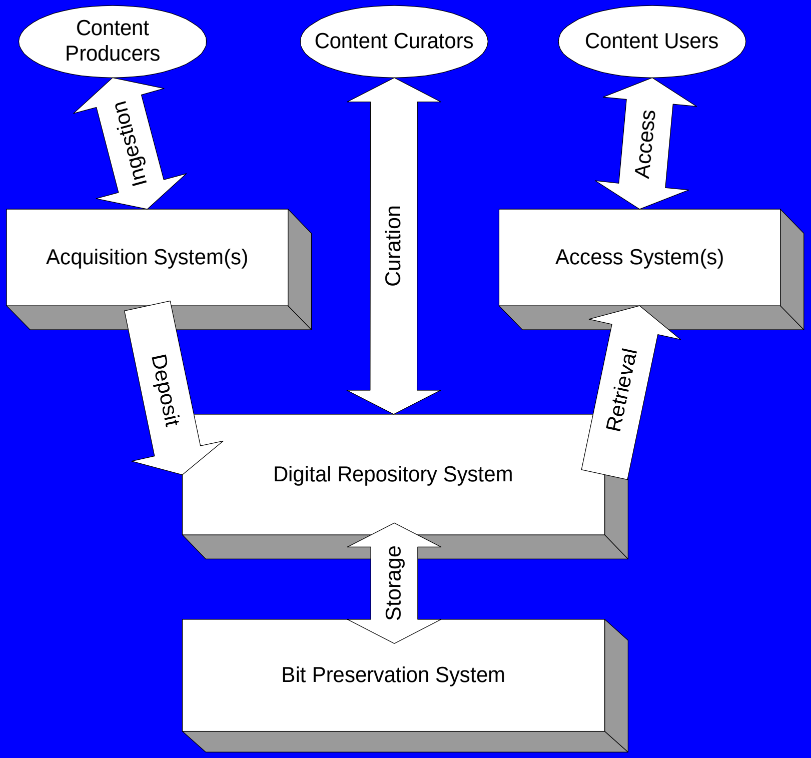
Outline

- Repository: definition, scope, services
- Development environment
- Case: National Digital Newspaper Program (NDNP)

Repository

- Software, hardware & processes that enable deposit, retrieval, & *preservation* of digital *objects*.

In the grand scheme of things!

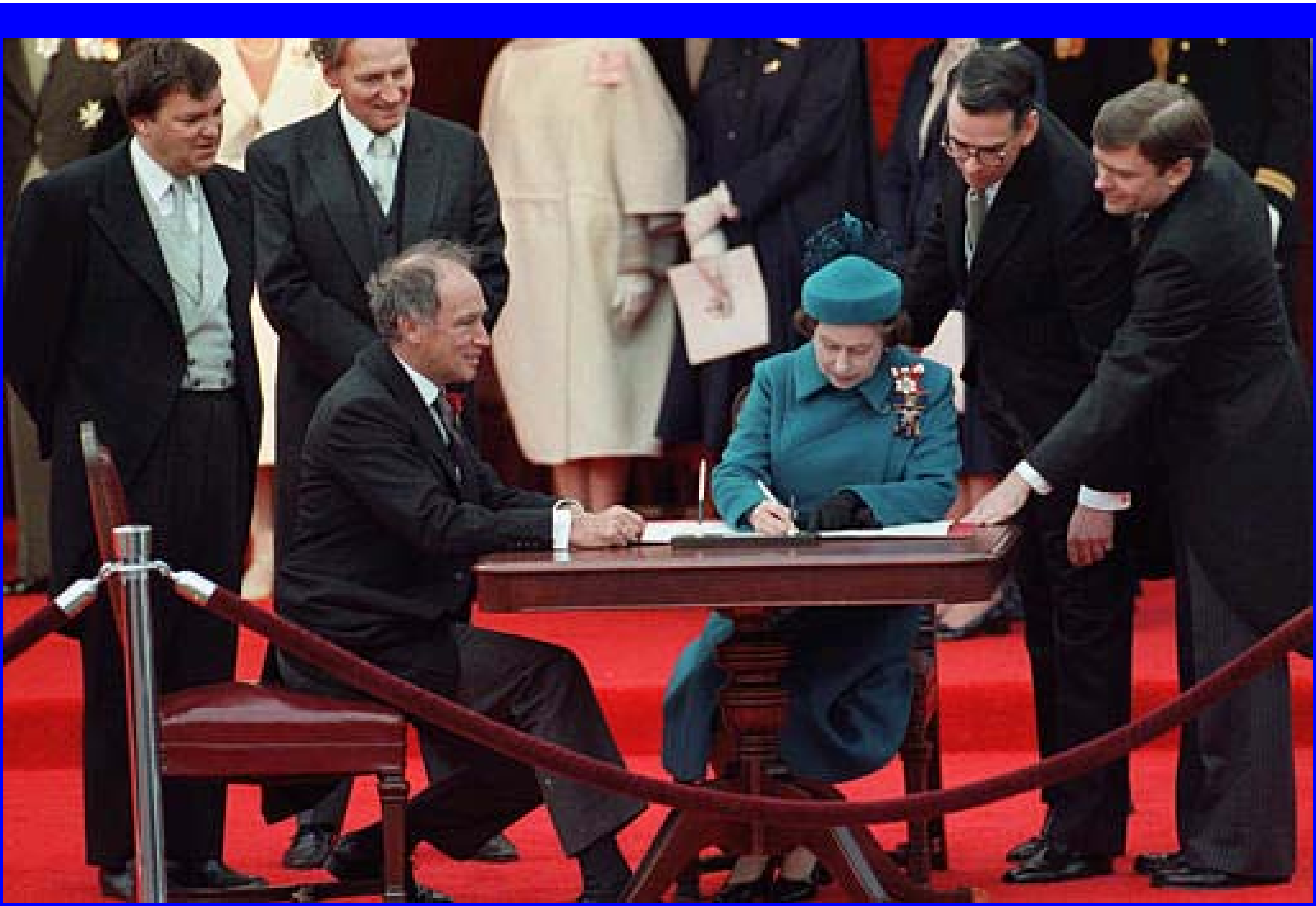


Bit preservation characteristics

- Files & directories
- Storage management
- Data replication & backup
- Checksuming
- Storage media refreshing

Repository characteristics

- Objects & relationships
- Content management
- Meta-data
- Context
- Migration (format)



What is this photograph showing?

Who took it?

Who owns it?

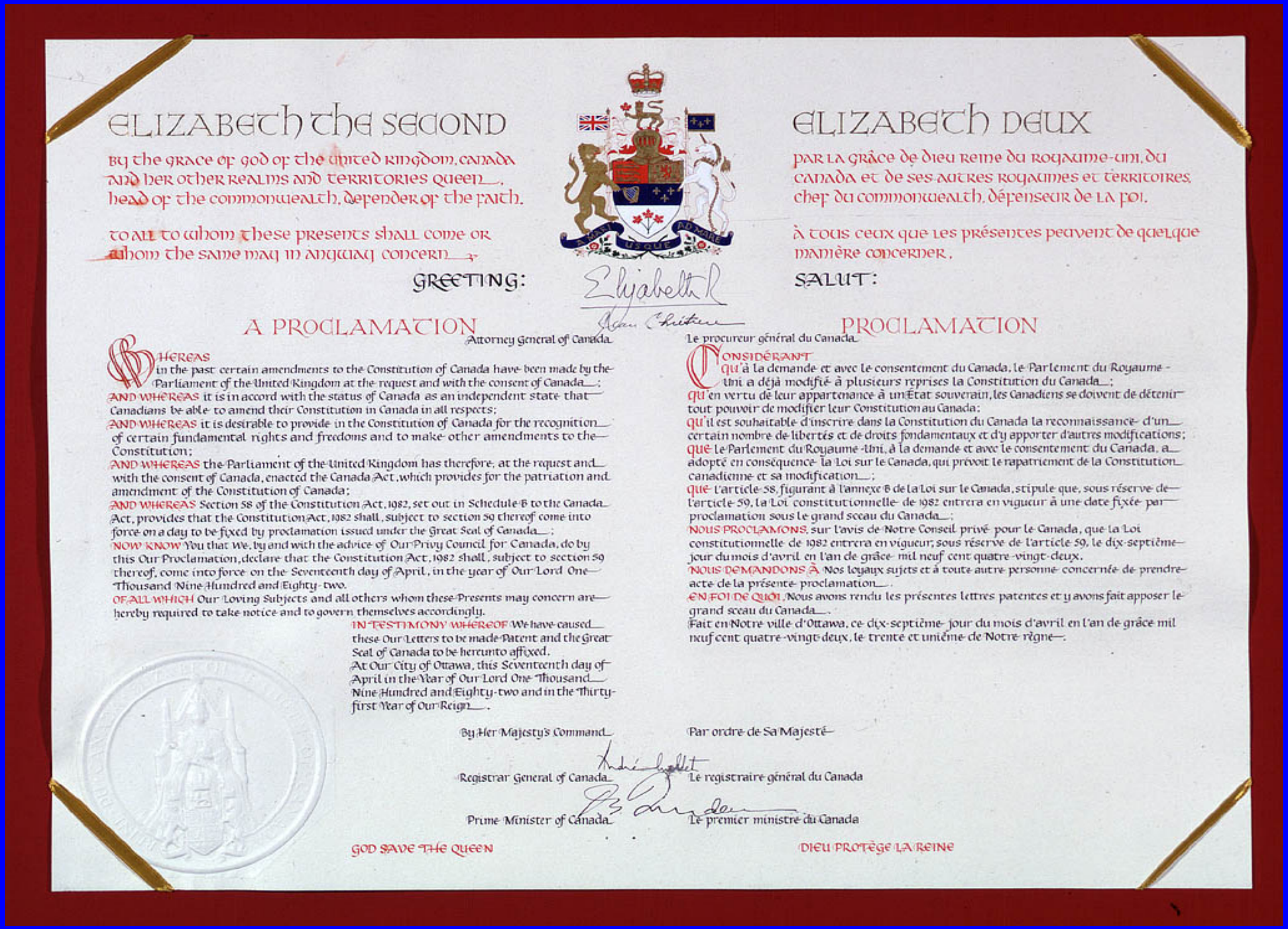
When was it taken?

Preservation

- Identity
- Integrity
  - Fixity
  - Completeness

Authenticity

- Understandability
  - Readability
  - Intelligibility



ELIZABETH THE SECOND

BY THE GRACE OF GOD OF THE UNITED KINGDOM, CANADA AND HER OTHER REALMS AND TERRITORIES QUEEN, HEAD OF THE COMMONWEALTH, DEFENDER OF THE FAITH.

TO ALL TO WHOM THESE PRESENTS SHALL COME OR WHOM THE SAME MAY IN ANYWAY CONCERN

GREETING:

Elizabeth R

A PROCLAMATION

Attorney General of Canada

WHEREAS in the past certain amendments to the Constitution of Canada have been made by the Parliament of the United Kingdom at the request and with the consent of Canada;

AND WHEREAS it is in accord with the status of Canada as an independent state that Canadians be able to amend their Constitution in Canada in all respects;

AND WHEREAS it is desirable to provide in the Constitution of Canada for the recognition of certain fundamental rights and freedoms and to make other amendments to the Constitution;

AND WHEREAS the Parliament of the United Kingdom has therefore, at the request and with the consent of Canada, enacted the Canada Act, which provides for the patriation and amendment of the Constitution of Canada;

AND WHEREAS Section 58 of the Constitution Act, 1982, set out in Schedule B to the Canada Act, provides that the Constitution Act, 1982 shall, subject to section 59 thereof come into force on a day to be fixed by proclamation issued under the Great Seal of Canada;

NOW KNOW You that We, by and with the advice of Our Privy Council for Canada, do by this Our Proclamation, declare that the Constitution Act, 1982 shall, subject to section 59 thereof, come into force on the Seventeenth day of April, in the year of Our Lord One Thousand Nine Hundred and Eighty-two.

OF ALL WHICH Our Loving Subjects and all others whom these Presents may concern are hereby required to take notice and to govern themselves accordingly.

IN TESTIMONY WHEREOF We have caused these Our Letters to be made Patent and the Great Seal of Canada to be hereunto affixed.
At Our City of Ottawa, this Seventeenth day of April in the Year of Our Lord One Thousand Nine Hundred and Eighty-two and in the Thirty-first Year of Our Reign.

By Her Majesty's Command

Registrar General of Canada

Prime Minister of Canada

GOD SAVE THE QUEEN

ELIZABETH DEUX

PAR LA GRÂCE DE DIEU REINE DU ROYAUME-UNI, DU CANADA ET DE SES AUTRES ROYAUMES ET TERRITOIRES, CHEF DU COMMONWEALTH, DÉFENSEUR DE LA FOI.

À TOUS CEUX QUE LES PRÉSENTES PEUVENT DE QUELQUE MANIÈRE CONCERNER,

SALUT:

PROCLAMATION

Le procureur général du Canada

CONSIDÉRANT
qu'à la demande et avec le consentement du Canada, le Parlement du Royaume-Uni a déjà modifié à plusieurs reprises la Constitution du Canada;

qu'en vertu de leur appartenance à un État souverain, les Canadiens se doivent de détenir tout pouvoir de modifier leur Constitution au Canada;

qu'il est souhaitable d'inscrire dans la Constitution du Canada la reconnaissance d'un certain nombre de libertés et de droits fondamentaux et d'y apporter d'autres modifications;

que le Parlement du Royaume-Uni, à la demande et avec le consentement du Canada, a adopté en conséquence la Loi sur le Canada, qui prévoit le rapatriement de la Constitution canadienne et sa modification;

que l'article 58, figurant à l'annexe B de la Loi sur le Canada, stipule que, sous réserve de l'article 59, la Loi constitutionnelle de 1982 entrera en vigueur à une date fixée par proclamation sous le grand sceau du Canada;

NOUS PROCLAMONS, sur l'avis de Notre Conseil privé pour le Canada, que la Loi constitutionnelle de 1982 entrera en vigueur, sous réserve de l'article 59, le dix-septième jour du mois d'avril en l'an de grâce mil neuf cent quatre-vingt-deux.

NOUS DEMANDONS À Nos loyaux sujets et à toute autre personne concernée de prendre acte de la présente proclamation.

EN FOI DE QUOI, Nous avons rendu les présentes lettres patentes et y avons fait apposer le grand sceau du Canada.
Fait en Notre ville d'Ottawa, ce dix-septième jour du mois d'avril en l'an de grâce mil neuf cent quatre-vingt-deux, le trente et unième de Notre règne.

Par ordre de Sa Majesté

Le registraire général du Canada

Le premier ministre du Canada

DIEU PROTÈGE LA REINE

How is this document related to the previous photograph?

Content Model

- Objects
  - Identity
  - Intellectual content
  - Description (attributes)
  - Behavior
- Relationships
  - Identity
  - Definition

Services and functionalities

- Unique, persistent, global identification
- Object inventory & registration
- Object representation
- Continuous, bulk ingest (validation, tagging, registration)

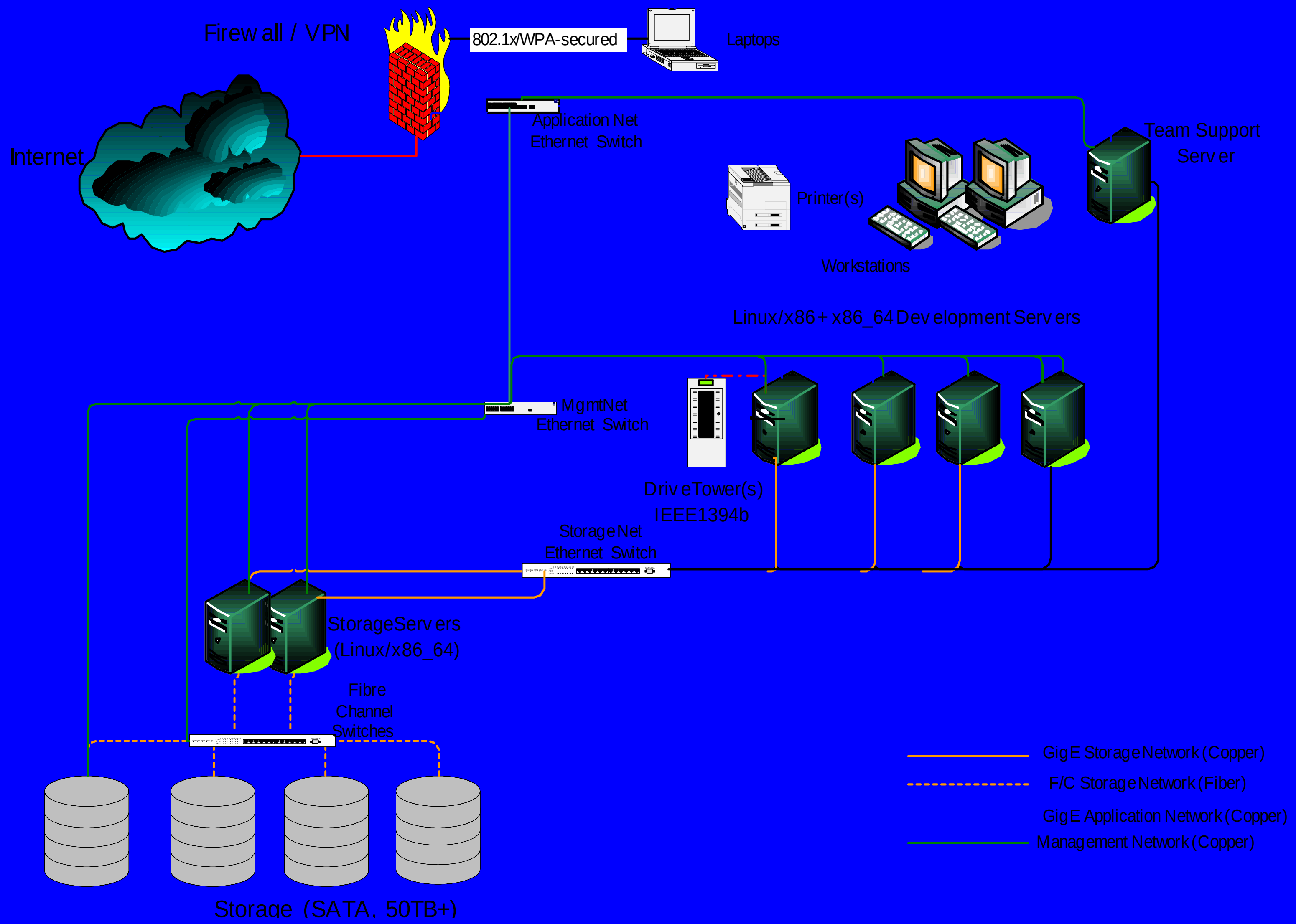
Services and functionalities

- Automated migration (w/ validation)
- Version management
- Rights management
- Meta-data management (updates, content association, preservation)
- Search and retrieval (content & metadata)

Repository Development Center (RDC)

- Architecture
  - – Commodity HW
  - – Inexpensive
  - – Scalable
  - – Configurable
- Open Source
  - – Content format
  - – System software
  - – Application software
  - – Interoperable

RDC Architecture



Case: NDNP

- NEH/LC collaborative program
  - NEH: Funds the program (“We the People” initiative)
  - Awardees: Select and convert
  - LC: Aggregates, preserves and serves
- Content:
  - Granularity: Newspaper page
  - Model: Page, Section, Issue, Title
  - Materialization:
    - OCR’d text
    - Page image

Case: NDNP

- Volume:
  - – 147,000 Newspaper titles
  - – 63 MB/Page
  - – 750,000 Pages
  - – Total: ~50 TB's
- Delivery to LC
  - – 500 Gb hard drives (approx. 8000 pages)

Current generation

- Formats
  - Page image: TIFF 6.0; JPEG 2000; PDF
  - OCR: XML, ALTO Schema
- Pre-Ingest:
  - Submission package tool
  - Validation tool
    - Visual
    - Automated
    - Authentication
    - Adding preservation metadata

Current generation

- Search and access
  - Full-text
  - Hit highlighting
- Repository prototype:
  - Ingest
    - Unique ID assignment
    - Automated file placement
  - Search
  - METS/MODS objects
  - Metadata

Next generation

- Delivery through Internet2
- Article-level granularity
- Object representation
- Preservation
  - Plans
  - Functions

Next generation

- Metadata
  - Support different standards
  - Preservation
- Scale
- Automated

Impact

- Modes of partnership
  - Funding
  - Production
  - Preservation
- Use & adoption of standards & best practices
- Open source architecture
- Large-scale preservation

More Information

- Website: http://www.loc.gov/ndnp
- Email:
  - babak@loc.gov
  - ndnptech@loc.gov